Durham University

Durham Research Online

Deposited in DRO:

26 February 2019

Version of attached file:

Accepted Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Ehren, M.C. and Jones, K. and Perryman, J. (2016) 'Side effects of school inspection : motivations and contexts for strategic responses.', in Methods and modalities of effective school inspections. , pp. 87-109. Accountability and educational improvement.

Further information on publisher's website:

https://doi.org/10.1007/978-3-319-31003-9_5

Publisher's copyright statement:

This is a post-peer-review, pre-copyedit version of Ehren, M. C. Jones, K., Perryman, J. (2016). Side Effects of School Inspection; Motivations and Contexts for Strategic Responses. In Methods and Modalities of Effective School Inspections. Ehren, M. C. Springer International Publishing. 87-109. The final authenticated version is available online at: https://doi.org/10.1007/978-3-319-31003-9_5

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the full DRO policy for further details.

Durham University Library, Stockton Road, Durham DH1 3LY, United Kingdom
Tel : +44 (0)191 334 3042 | Fax : +44 (0)191 334 2971
https://dro.dur.ac.uk

Chapter 5: Side effects of school inspection; motivations and contexts for strategic responses

Melanie Ehren, Karen Jones and Jane Perryman

This chapter introduces three categories of unintended consequences from school inspections: 1) intended strategic behaviour where schools manipulate the inspection assessment through window dressing, misrepresentation or gaming, 2) unintended strategic behaviour when schools narrow their educational practices as a result of the behaviour of the assessor and/or by the method of working used for the assessment, and 3) other types of consequences, such as stress, anxiety and increased workload. As many inspection systems use standardized student achievement tests to evaluate school output, a fourth category on unintended responses to high stakes testing will also be introduced.
The results from a recent systematic literature review will be used to provide evidence of responses in each of the four categories. The review shows that most studies present examples from England and previous case study work from Perryman (2006) will therefore offer more in-depth views of how an English school responds strategically to school inspections. The final section of the chapter provides explanations of the conditions under which such responses may occur.

5.1. Introduction

Since the 1990s it has been widely documented in literature that accountability systems, including school inspections, bring with them unintended consequences. These unintended consequences are often negative and have the potential to undo the intended positive effects. Despite the vast sums of public money spent on inspection throughout Europe and beyond, very little evidence exists as to how prevalent these potential side-effects are. This chapter will outline different types of unintended consequences, such as intended strategic behaviours where schools manipulate the inspection assessment through window dressing, misrepresentation or gaming, and unintended strategic behavior when schools narrow their educational practices as a result of the behavior of the school inspector and/or by the frameworks and working methods used to inspect schools (De Wolf and Janssens, 2007). Such frameworks and methods often include an examination of a school's aggregated test results and league tables. Where this is the case strategic responses will likely also include examples of teaching to the test as found in test-based accountability systems.
This chapter will draw on the previously described systematic literature review to provide an overview of the prevalence of the intended and unintended strategic behaviours and teaching to the test of schools in response to school inspections. Evidence from 49 studies from England (33 studies), the Netherlands (2 studies), Turkey (2 studies), Flanders (2 studies), Hong Kong (2 studies), Ireland (1 study), France (1 study), New Zealand (1 study), and Germany (1 study) will be used to give examples of the three types of responses across Europe, analysing differences in strategic responses of teachers compared to principals, and analysing the characteristics of inspection systems (e.g. perceived pressure and high stakes) that motivate such responses.

5.2. Categories and types of unintended consequences of school inspection

Many studies have reported how the implementation of rules and regulations, the monitoring of the implementation of such rules and regulations and performance in general may lead to behaviour which was not intended. As Jones et al (in prep, p.4) explain, this was termed "unintended consequences" by Merton (1936) and his heavily cited article suggested that these unintended consequences can be positive or negative. The expansion and study of performance measurement schemes in the social sciences in the 1970s led to the discovery

and formulation of Cambell's Law, which has become well cited in studies about unintended effects of educational accountability:

> *The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.*
> Campbell (1976, 49)

How performance measures can corrupt behaviour was explained in more detail in Smith's (1995) seminal paper about the unintended consequences of standardized measures of performance. He explains how most performance indicator schemes will fail unless consideration is given to their deficiencies. In his paper he provides examples from a number of different sectors on how performance indicators can lead to tunnel vision, misrepresentation and a range of other behaviours. A number of authors have used his examples to explain similar strategies in schools responding to league tables and the publication of high stakes inspection assessments (e.g. Fitz-Gibbon, 1997). More recently, Smith's overview of strategies have been used by De Wolf and Janssens (2007) to summarize the research into side effects of performance standards and school inspections, categorizing his strategies into three distinct categories of intended and unintended strategic behaviours and 'other' unintended consequences. These three categories guided our literature review and are used below to summarize our findings.

5.2.1 Intended strategic behaviour

The first category of unintended consequences is termed *'intended strategic behaviour'* and refers to principals' and teachers' responses that are intentionally taken to improve the school's inspection assessment. Responses include window dressing, fraud, gaming and misrepresentation.

Window dressing occurs when schools implement procedures and protocols that have no effect on primary processes in the school, but are implemented to be assessed more positively. Schools are 'brushed up' to receive a more positive assessment. They can use several methods that vary in fairness and lawfulness to do so, such as fraud, gaming and misrepresentation. Fraud occurs when schools falsify numbers or records (such as test scores or lesson plans) used in school inspections to assess output or educational processes of/in schools. Misrepresentation occurs when schools manipulate behaviour they have to report on. Examples are excluding low performing students from exams that are used to assess schools as these students may lower the average test scores of schools. Gaming refers to schools manipulating actual behaviour.

Most studies provide examples of intended strategic responses of teachers and head teachers to Ofsted inspection in England, explaining how teachers plan artificial lessons to please Ofsted during inspection visits, gearing teaching towards what inspectors are going to value and measure as outlined in the inspection framework documentation and 'teach to inspection' (Webb et al, 1998; Kogan and Brunel University, 1999; Hall and Noyes, 2007; Richards, 2014; Hardy, 2012). Kogan and Brunel University (1999) and Perryman's (2009) study also shows how teachers and managers 'perform' during an inspection and put up a show. Chapman (2001) and Brimblecombe (2000) explain how the lessons taught during an inspection visit are more highly prepared than normal, and how teachers plan and deliver 'safer' and more teacher-led classes to avoid the possibility of loss of control during the inspection observation. Fitz-Gibbon and Stephenson-Forster (1999) found that 81% of principals surveyed claimed that inspectors did not see the school as it normally is, where in Brimblecombe et al's (1996) study a third of surveyed teachers suggested the inspector did not see a typical lesson. Roberts-Holmes (2014) also explains how teachers in England game the system when they adjust profiles of students such that a decent number of them attained good levels of development.

A recent EU-study by Ehren et al (2013; see also Jones et al, in prep) also asked principals in England, the Netherlands, Ireland, Sweden and Austria about potential manipulation of inspection data, and about putting procedures and protocols in writing in preparation for inspection visits. The findings from this study suggest that these intended strategic responses are more prevalent across all the five countries than unintended strategic responses. Principals in all the five countries, except for the Netherlands agree that school inspections are about putting protocols and procedures in writing in preparation for the inspection visit. Principals in all the countries, except for Ireland, however disagree with presenting a more positive picture of the school to the Inspectorate.

5.2.2 Unintended strategic behaviour

Unintended strategic behaviour is a second category of unintended consequences. According to De Wolf and Janssens (2007) this category refers to the unintended influencing of behaviour by the assessor and/or by the method of working used for the assessment. In effect this means a (usually unintended) one-sided emphasis on the elements that are assessed, such as described in the inspection framework and protocols and guidelines inspectors use during school visits. De Wolf and Janssens (2007) describe a number of examples of unintended strategic behaviour, such as 'tunnel vision' where schools emphasize phenomena that are quantified in the performance measurement scheme, at the expense of unquantifiable aspects of performance. Schools focus for example on programming a large number of lesson hours instead of trying to improve the quality of lessons offered.
Suboptimization is another example of unintended strategic behaviour. Local objectives are pursued by schools, at the expense of the objectives of the school as a whole, such as when teachers and principals focus the curriculum and teaching on mathematics and literacy when those are inspected, ignoring other subject areas that are important for students' success in later (school) life. Myopia, a third example, includes schools pursuing short term targets (for example improving test scores by means of redirecting students to easier subjects) at the expense of legitimate long term objectives (improving student achievement in difficult subjects). Schools aim at success that can be established very quickly, instead of long-term school improvement. Ossification, or organizational paralysis, is a fourth type of unintended strategic behaviour which occurs when schools refrain from innovating and ignore changes and threats, because innovative arrangements are not rewarded in the inspection framework. Schools are expected to suffer from ossification when performance measurement schemes are used rigidly. Measure fixation is a last example of unintended strategic behaviour and refers to schools that focus on measures of success rather than the underlying objective. Schools implement, for example, self-evaluation instruments to score positively on inspection indicators used for measuring quality assurance, instead of implementing such instruments to improve the quality of their education.

Our systematic literature review provides evidence of unintended strategic behaviour of teachers and principals across a number of countries, but particularly England. A recent EU-study (see Jones et al, in prep) shows how principals in England scored significantly higher on items around unintended strategic responses, particularly on items about refocusing and narrowing the curriculum and teaching and learning strategies to meet inspection criteria, compared to their peers in the Netherlands, Ireland, Sweden, and Austria.
School inspections also seem to slow down school development when schools suspend school activities when time is needed to prepare for inspections or had less energy for change after an intense visit (Ouston et al, 1997; Hopkins et al, 1999; Plowright, 2007; Berry, 2012). Nees (2006) reports of concerns about the sustainability of improvements made in six primary schools in Wellington (New Zealand) in response to external review, including recruitment challenges and barriers to learning among their students. Hopkins et al (1999) also explain how relationships within schools suffer from the pressure to prepare and undergo an Ofsted inspection (Hopkins et al, 1999), and how the normal running of the school is interrupted

during an inspection visit when inspectors for example influence the behaviour and contributions of pupils in lessons (Dean, 1995; Gray and Gardner, 1999).

An OECD report (2013) refers to a study from West et al (2011) who explain the shift from an improvement culture to one of compliance to inspection criteria as a result of the publication of league tables, school competition and the risk of sanctions for failure in inspections. According to Braun et al (2010), external pressures, such as from Ofsted, reinforce stereotypes about students and learning; they describe a specific example of how beliefs about female students needing to involve themselves actively in lessons in order to learn and achieve were reinforced.
Day and Gu (2010), Lupton et al (2012) and Chapman (2002) clarify how schools who are not improving loose local authority confidence, are under increased scrutiny by both the local authority and the inspectorate, loose professional autonomy and motivation and move towards a 'low-risk culture'. The two schools in Lupton et al's (2012) case study felt encouraged to move to narrower forms of pedagogy and 'teaching to the test'. The combination of the schools' low Ofsted rating and position in the league tables, their disadvantaged circumstances and the consequent ways in which both teacher and pupil identity are constructed and institutionalised were seen as the main causes of such unintended strategic responses.

5.2.3 Other unintended consequences

Our systematic literature review suggests a range of other unintended consequences from school inspections. A large number of studies from England describe the intense anxiety, fear, high workload to prepare paperwork and provide data, stress and demoralizing of staff associated with Ofsted inspections, often leading to time off for illness after inspections, particularly in schools who were placed in special measures (Jeffrey and Woods, 1996; Sebba et al, 1996; Webb et al, 1998; Gray and Gardner, 1999; Scanlon, 1999; Lee and Davies, 2000; Chapman, 2000, 2002; Case, 2000; Follows, 2001; Keeble-Allen, 2004; Hardy, 2012; Berry, 2012). The special measures label was punitive and stigmatising for schools and the sense of being permanently under a disciplinary regime can lead to fear, anger, disaffection and a loss of power and control, according to Perryman (2002, 2007). Jeffrey and Woods (1996) explain how teachers feel a 'loss of pedagogic values' – described as a grieving process, teaching being reduced to numbers and percentages that were 'satisfactory', and people feeling persecuted and guilty through the exercise of bureaucratic controls. According to Jeffrey and Woods (1996), inspections led teachers to lose confidence in their professional role and to redefine their profession from a moral profession to a more instrumental one, aimed at achieving high test scores.

Stress and anxiety is in most schools highest in preparation for the inspection. In Kogan and Brunel University's study (1999), teachers explain how they felt the inspection itself was less traumatic than expected. Brimblecombe et al (1995) however describe how teachers also experience high levels of stress during an inspection visit when inspectors behave in an 'inspectoral' or punitive manner (focusing on negative issues in the school), and this can, according to Berry (2012), Bates (2013) and Courtney and Steven (2013) undermine confidence and commitment, increase cynicism and resentment about the inspection process, and reduce overall trust within the system. Findings from Courtney and Steven's qualitative study (2013) suggest that outstanding headteachers are less likely to take a job in a deprived area due to worse job security and greater difficulty in achieving a high grade and felt that there was a 'climate of fear' created by Ofsted's inspection framework. Some of the anxiety may be reduced through pre-visit contact between school and inspector, particularly when inspectors are reassuring (Dean, 1995).

Increased workload is also mentioned in many studies as an unintended consequence of Ofsted school inspections (e.g. Jeffrey and Woods, 1996; Keeble-Allen, 2004; Lee and Davies, 2000; Hardy, 2012; Jones et al, 2015). Hall and Noyes (2007) describe how the

requirements for self-evaluation increased the workload when preparing for an inspection, whereas Kogan and Brunel University (1999) and Scanlon (1999) quote school staff reflecting on their preparation for an Ofsted inspection visit as being a very bureaucratic exercise.

Lupton et al (2012), Courtney (2012) and Brookes (2008) also explain how inspection outcomes have consequences for headteachers' careers who often lose their job as a result of a low Ofsted grade, contributing to poor retention and recruitment of school leaders.

Similar side effects on workload and stress levels have been reported in Hong Kong, particularly in schools from 'weak settings' who had problems with recruiting and retaining suitable staff (MacBeath, 2008; Wong and Li, 2010). MacBeath (2008) and Chapman (2002) describe how school leaders in England and Hong Kong have an important role in mediating levels of stress and high workload among/of staff by supporting external review as an opportunity for improvement, instead as a threat.
Ehren and Shackleton (2014) additionally found unintended consequences in secondary schools in the Netherlands who were judged to be weak or very weak by the Dutch Inspectorate. These schools saw a decline in student satisfaction and in student numbers following the inspection.

5.2.4 Unintended consequences of high stakes testing

A different strand of the literature covers the side effects from high stakes testing and league tables, such as teaching to the test. As many school inspection systems include student achievement data in their data collection to assess school quality, research and the evidence base on high stakes testing and test-based accountability is also relevant in thinking about unintended consequences from school inspections.
Research on high-stakes testing has particularly been carried out in the United States since the introduction of test-based accountability in the 1990s culminating in the 'No Child Left Behind' Act introduced in 2001. The bill requires states to judge the performance of schools based on annual test scores. As a result of these changes there have been a number of studies from the US focussing on the effects and side effects of high stakes testing. Several authors have tried to distinguish teacher practices to prepare students for testing that are ethical, legitimate, have beneficial effects on learning and lead to valid increases in test scores from less beneficial activities.

Koretz, McCaffrey and Hamilton (2001) for example propose a continuum of teacher responses to high stakes testing in which they distinguish responses that are likely to be positive, ambiguous or negative in promoting student learning, and in leading to (in)valid test scores. As the connection between specific teacher responses to testing and learning gains of students has not been studied, this list is hypothetical.
Positive responses are, according to these authors, those responses that are expected to lead to beneficial effects on learning and to valid increases in test scores. Examples of such responses are when teachers provide more instructional time, cover more material or teach more effectively. Haladyna et al (1991) also consider increasing student motivation to do well on the test, checking answer sheets to ensure they were completed properly and teaching test-taking skills ethical and positive responses.

Koretz, McCaffrey and Hamilton (2001), Stecher (2002) and Booher-Jennings (2005) also identify responses of teachers to high stakes tests whose impact is ambiguous; the impact depends on the specific circumstances. Examples of ambiguous responses are:

- reallocating instructional resources (classroom time or students' study time) within and across subjects to emphasize topics covered by the test instead of content that receives little or no emphasis on the test
- coaching students to do better by focusing instruction on incidental aspects of the test

- aligning instruction with standards to give material and curriculum content that is consistent with standards more emphasis
- targeting instructional resources to students close to a cut-point set in the accountability system to improve the school's overall score on the accountability measures (educational triage)

According to Stecher (2002), reallocation, alignment and coaching may have positive consequences when they focus on important aspects of the domain the test is designed to measure or specific skills that help students demonstrate their actual achievement. Students will be offered more time and resources to learn these domains and skills. Also, familiarizing students with the format and other aspects of a test can increase the validity of scores when certain mistakes on tests that are made because students don't understand test instructions or question formats, are prevented.
Reallocation and alignment have positive consequences for tested subjects that are emphasized in the curriculum and instruction, but negative consequences for important aspects of the curriculum that are not represented in the test and are now being ignored (for example because they are difficult to assess in a multiple choice test format, such as listening and speaking skills) (Stecher, 2002).
According to Stecher (2002), coaching can also lead to negative consequences when it improves test performance by focusing on features of the test that are incidental to the domain the test is supposed to measure. Because these features are incidental, learning about them does not produce real improvement in students' knowledge of the domain. In this case, teachers instruct pupils to do well on a test but fail to teach them to transfer this knowledge to different types of problems or even different formats of test items.

Negative responses are those that are considered to be unethical and harmful for student learning and will lead to invalid increases in test scores. Examples of such responses are, according to Stecher (2002), when teachers respond to high-stakes testing by cheating and distorting data used to measure the school's status on the accountability measures or when using instruction time to ineffectively drill and practice students for the test. Jacob and Levitt (2003) revealed 4-5% of the classrooms in their US-based study cheating each year. Teachers may do so by prompting students with the right answer during a test, providing the actual test items in advance, providing hints during test administration, making changes to answer sheets before scoring or leaving pertinent materials in view during the testing session. These responses lead to harmful consequences for student learning or inflated test scores.
Figlio and Getzler (2002) and Cullen and Reback (2006) also describe how schools at risk of failing improve their state-assigned grade or classification by taking their poorest performing students out of the testing pool. This type of response is usually referred to as 'reshaping the test pool'. Schools may do so by classifying (regular) students into the 'special education' or 'limited English proficient' categories that may be exempted from taking the test (Jacob, 2005; Cullen and Reback, 2006; Figlio and Getzler, 2006). Other methods used are retaining low-scoring students in grades below those in which the test is administered, directing students to lower educational tracks, allowing an increase in absences on test days, granting exemptions from testing by parents of low achieving students and increasing dropout rates of low achieving students.
The table below provides a summary of these responses.

Table 1. Classification of teachers' responses to high stakes testing

Positive responses
- Providing more instructional time
- Working harder to cover more material and content
- Working more effectively (achievement orientation, quality of instruction)
Ambiguous responses

- Reallocation of instructional resources (time and material)
- Coaching
- Aligning instruction with standards
- Targeting instruction to students close to accountability target (educational triage)
Negative responses
- Using instruction time to drill and practice for the test
- Reshaping the test pool
- Cheating

There is currently limited evidence on the extent to which these responses to high stakes testing also occur in inspection systems that use standardized tests to evaluate schools. Available research of Ehren and Swanborn (2012) on schools' responses to the use of student achievement data by the Dutch Inspectorate of Education suggests some reshaping of the test pool and questionable test administration practices (e.g. clarifying test questions to prompting students with the correct answer) in the Netherlands, but as there were no key differences between high and low performing schools it is possible this was due to other motives rather than aiming to improve inspection evaluations.

Sturman (2003) studied survey data from primary schools in England and found some evidence on teaching to the test and showed that test preparation occupied a large proportion of time. However Sturman argues that these practices can have beneficial effects as well as negative effects leading to score inflation. Tymms (2004) also suggests that test score rises in English primary schools were partly due to teaching to the test practices. Wiggins and Tymms (2002) use survey data to compare primary schools in England (where league tables of examination results are published) with Scotland (where no results are published). They find large differences between Scottish and English schools, with English schools reporting more concentration on performance targets at the expense of other important objectives, a greater 'narrowing effect' on the curriculum due to testing and a greater focus on 'borderline' students (those close to the border for national target levels) at the expense of other students. Similar findings have been reported by Kogan and Brunel University (1999), Hall and Noyes (2007), Hardy (2012), Lupton and Hempel-Jorgensen (2012), and Roberts-Holmes (2014) who talk about the narrowing of teaching and the curriculum to exams and to focus on progress in literacy and numeracy, concerns about exam results constraining pedagogical practice and 'substantive student learning in general', and narrowing the understanding of successful pupil outcomes to those that are readily measurable by testing.

Roberts-Holmes (2014) describes how teachers try to game the system by adjusting profiles of students such that a decent number of them attain a 'good level of development' to prevent intensive scrutiny by Ofsted. Additionally, Hardy (2012) discusses a tendency for teachers to focus on students near pass/fail borderlines and to focus on practices to improve results rather than student learning overall. A large quantitative study by Hussain (2012) however does not corroborate these results as his examination of schools' improvement in relation to prior attainment showed improvement for all the students in the schools, suggesting that schools do not target borderline students or fail to enter students less likely to perform well.

Ehren and Jones, in a comparative EU-study on the impact of school inspections (see Ehren et al, 2013) surveyed teachers from different grades in primary and secondary education in the Netherlands and England about their responses to standardized tests used by the Inspectorate of Education in both countries. As table .. shows, teachers in both countries report little activities to prepare students for the test; they disagree on all the statements asking them about coaching students and using drill and practice to prepare students for standardized tests, or reallocating their curriculum and instruction to tested subjects. There are however some significant differences between teachers in England and the Netherlands with teachers in

England generally reporting higher scores for coaching students to do well on the test and using drill and practice to prepare them for the test. Teachers in the Netherlands however report more activities to align their curriculum and instruction to the test, particularly in reviewing the test to decide on which topics to teach.

Table .. Teachers' responses to high stakes testing (combined teacher data, Y1, 2, 3)

		N	Mean	SD	Mean difference (for significant differences between country means)
Teaching to the test (14 items, α = 0.84)	England	768	2.48	.57	
	Netherlands	696	2.43	.75	
Reallocation (2 items, α = 0.64)	England	655	2.18	.90	-0.39**
	Netherlands	680	2.57	1.10	
I reviewed what was on last year's test when designing my lesson plan/pacing calendar for this year	England	655	2.16	.950	-.410**
	Netherlands	680	2.57	1.095	
I adjust my instructional plans based on the test/exam performance of the class I had last year	England	185	2.43	.936	
	Netherlands	0[a]			
I teach topics that are not on the test/exam after	England	646	2.29	.982	.140*
	Netherlands	687	2.15	1.308	
Coaching (5 items, α = 0.82)	England	663	2.73	.71	0.18**
	Netherlands	696	2.55	0.94	
Teaching test-taking skills, such as pacing/timing, strategies for answering multiple-choice questions, eliminating wrong m.c. options, filling in answer sheets	England	656	2.71	1.046	-.146*
	Netherlands	626	2.86	1.212	
Explaining questions from the test that was administered last year	England	652	2.57	.963	.454**
	Netherlands	599	2.12	1.106	
Discussing responses to the test items	England	654	2.83	.947	.827**
	Netherlands	608	2.00	1.080	
I emphasize particular styles and formats of problems in the test/exam in instruction (e.g. using particular styles of graphs; using specific key phrases)	England	654	2.85	0.91	.201**
	Netherlands	687	2.65	1.06	
Within 1 month of testing, I provide a "refresher" on content and/or skill areas that specifically match those on the test/exam.	England	651	2.73	.986	
	Netherlands	693	2.74	1.261	
Drill and practice (4 items, α = 0.70)	England	660	2.36	0.69	0.15**
	Netherlands	694	2.21	.96	
Having students practice old tests	England	452	2.56	.828	.591**
	Netherlands	587	1.97	1.068	
Providing practice on questions from the test that was administered last year	England	652	2.60	.981	
	Netherlands	609	1.61	.915	
I use m.c. questions from previous tests in	England	647	1.92	.990	-.383**

my classroom assessments.	Netherlands	687	2.31	1.403	
Within 1 month of testing, I use practice exercises/tests with multiple-choice questions and language similar to that found on the test/exam.	England	651	2.49	1.069	-.153*

Note: technical report for description of sample, data collection and analyses can be found on www.schoolinspections.eu

5.3. An example from England

As many studies of unintended consequences from school inspections are from England, this section presents a case study from Perryman (2006) as an example of some of the responses described before. This particular case study was conducted under the auspices of the 2003 Ofsted school inspection framework, when schools had long notice periods to prepare for inspections.

In Perryman's (2006) research in Northgate, the case study school was in special measures. At Northgate, working under special measures, teachers found that the best way to get out of this category was to act as if they were being inspected all the time, in order to train themselves and pupils into expected modes of behaviour, and so that the arrival of an inspector would be easier to deal with, and part of the routine. Ball's phrase 'an organisation for 'the gaze' and for the avoidance of 'the gaze'' (1997, p.332) seems very appropriate for Northgate, which, during special measures and before and during its Ofsted inspection, seemed like an organisation existing purely for the purposes of passing an inspection.

Documentation was used to both inculcate and demonstrate a discourse of effectiveness. Ball (2003, p.8) remarks that under inspection 'what is produced is a spectacle or what we might see as an 'enacted fantasy' which is there to be seen and judged'. He goes on to say that 'the heart of the educational project is gouged out and left empty. Authenticity is replaced by plasticity'. Under an inspection regime, a school's documentation becomes part of the surveillance. This is not unusual, as Duffy (1999, p.110) notes:

> Some of the documents generated by a school for an inspection may have the aim of giving the best possible impression to the inspectors, and the school might not be so prolific in its production of policy statements or so up to date in its handbook if the inspection was not imminent.

In schools this can be seen in constant clarifications of policies and procedures, in departmental handbooks, school and departmental action plans. Northgate was reinspected within two years of coming out of special measures, as is necessary in such circumstances. The school had learned its lessons well. As was usual under the 2003 Inspection Framework, the school had eight weeks' notice. Documentation was prepared, schemes of work and lesson-plans revamped and the full performative environment recreated. Extensive documentation was produced at department level.

In addition to department documentation, for the inspection the school had to produce a self-evaluation document called the S4, which under the 2003 Ofsted model, was very important. Ofsted would use the S4 to plan their visit, as the school would identify strengths and weaknesses. This was not a straightforward process, and would be the first time the game metaphor was made explicit:

> *A visiting governor, with Ofsted experience, made it very clear to the senior team that they had to play a clever game. Strengths had to be identified, but not overplayed; weaknesses discussed, but in the light of planned improvements. If Ofsted agreed with the self-evaluation, it was implied, then all would be well.*

In the 2005 framework the S4 was replaced by the even more stringent SEF (self-evaluation form) which served as the main document that Ofsted used to prepare its inspection of the school. The problem with self-evaluation documents produced for evaluation is that for some schools an honest warts-and-all approach is simply not possible. As far as some schools are concerned, over-emphasise strengths and they could be criticised for complacency with a management team unable to plan for progress, but identify too many weaknesses and there is a risk of giving a skewed picture which may influence the judgement of the inspectors negatively. Mel describes similar dilemmas with her subject documentation:

> *I'm worried that if I put that in the Department Review and say 'I'm going to do this, I'm going to do that' there will be too much 'I'm going to' rather than I'm there doing it* (Mel, middle manager).

Was Mel to admit she hadn't started many of her well-intentioned plans, or pretend everything was in hand? This approach does have implications for an institution's improvement, because if it disguises serious faults in order to avoid going into special measures, then it will not receive the support it may need.

Fabricating the stage

Another finding from Perryman's case study was the great deal of preparation done for the inspection. For example, before the Ofsted inspection, the school had a special in-service training day, specifically set aside for departments to work together on their lesson plans for Ofsted. Displays were created and erected. The Registered Inspector had requested work to be collated for six pupils per year into subject boxes; two from pupils of higher ability, two medium and two lower. There was frantic activity after school as departments selected their book samples. They were playing the game of selecting two 'low achievers'' books that 'aren't too low'. As someone remarked, 'If someone is 'low', won't their book be really crap?' (field notes).

In the week before the inspection, Lola, a head of department, wearily listed the extra work she'd had to do, and expressed the desire to just get it over with:

> *We've had to be observed and jump through the hoops for the observations. We've had to put a lot of unnecessary paperwork and things in place that were in place, but it's now decided that it's got to be done in a certain format, because everybody's got to use the same format.* (Lola, middle manager)

Everyone had to use exactly the same lesson plan format, schemes of work were written to a rigorous formula. There was no room for deviation.

Playing the game

At Northgate, preparing the stage went far beyond just the physical environment. There was a real sense from middle and senior managers that they were playing a game. This not only involved jumping through the prescribed hoops, teaching lessons in the correct manner, presenting all the correct documentation etc, but also suppressing negative thoughts and comments – and even hiding some pupils. Apart from lesson observations, the inspectors would be speaking to around 70 pupils and stressed that it was important to have quality time with them. The meeting schedule for teachers was organised. Meetings were rehearsed and these findings are echoed by Grace (Jeffrey and Woods, 1998, p.155) 'We practised ensuring that we presented a consensus for any interviews we had. It was very helpful. I want them to say that the Senior Management Team has a shared clear view'.

It was in the stage-managing of the morning briefing that the performance seemed really blatant. Morning briefings occurred at Northgate before the beginning of every school day, as the whole staff met to hear and give announcements (such as timetable changes, staff absences etc). The inspectors would expect to attend briefings, and Perryman's field notes remark on the way in which there was a rehearsal of how the first morning meeting in the

presence of the inspectors would go, and its subsequent success. During the inspection week, she wrote:

> *The staff briefing goes as rehearsed. The room is very crowded, with the addition of 13 inspectors. Lots of 'showy' things said, most of which were already announced on Thursday – football scores, a cricket award and a Year 9 trip to the Globe Theatre. One head of department said 'it's all so bloody false. I'd like to know what people are doing this week that they're not normally doing'.*

Resistance and cynicism

Teachers cannot refuse to be inspected, but can resist it in how they perform and stage manage what is seen, and in their grudging response to inspectors' judgements, with 'failing' viewed as a political construct, and 'success' greeted with cynical snorts.

In one example of resistance, some of the performance went as far as actually fooling the inspectors. As Helen reflected a year later:

> *I think the problem is that it is too easy to fool Ofsted. I think it is very easy to present them with superficialities in terms of paperwork that they are impressed by, and then when they go away you know that this is just a facade and therefore you have no respect for the whole process and that is how I feel about it* (Helen, middle manager).

Similarly, a school middle manager told Stoll and Fink (1996, p.5), 'I thought there were a lot of issues, both good and bad, that the Ofsted report hadn't even touched on here… it didn't focus very deeply on issues which affected the school'. This is an important issue as, if schools are hiding their weaknesses (and thus their real development needs) from inspectors, then real progress and improvement will be severely hampered. Plowright (2007, p.384) also found that the school he researched covered up its real problems, one head of department complaining 'each time…they paper over the cracks and it looks fantastic on the report. Whereas you only have to go a little bit deeper and there are real problems'.

Continuing the theme of fooling the inspectors, was the issue of the rather sinister sounding 'redirected pupils'. At Northgate around 20 of the most troublesome students were being sent off on various activities during Ofsted week, some on a residential trip to an outdoor activity centre, others on a programme of educational day trips with their learning mentors, some of which would lead to Duke of Edinburgh Awards. This is not an unusual strategy adopted by schools during inspection. A correspondent on the *Times Educational Supplement* forum comments:

> The pupils were well behaved which the inspector commented on. This is perhaps not surprising, as due to a reciprocal agreement with heads of other local schools, many of the most challenging pupils were on 'step out' visits for the duration. A large number of others ended up on temporary exclusions. ('halfmeltedsnowman' in Duffy, M. 2005)

MacBeath (2004) concurs, noting of one English secondary school under inspection that 'troubled students were sent away to an outdoor pursuits centre to partake in a week long alternative education system'.

Replacing teachers or ensuring the less capable were not present during the week of the inspection has also been reported. An interesting example was given by an AST (advanced skills teacher) on the TES forum. The posting is reproduced below.

> I currently work as an Advanced Skills Teacher (AST) in the London area. Having started my secondment to a struggling school in North London; I had been there for one day a week for four weeks; I was apprehended at the end of my Year 11 parents evening by my Principal. He asked me if I could go into my secondment school the following morning and teach their Science NQTs timetable as the school was being Ofsteded and the SLT had asked her to stay at home. Firstly, I must state I am a Chemist not a Physicist and this lady taught primarily Physics; secondly it was 8.30 at night; thirdly the school expected me to plan all the lessons outright. I

> foolishly agreed to go in to support but was not happy about the situation and had been assured nobody would visit my classroom. On arrival I was met by the school's lead deputy who handed me a wedge of data and escorted me to the conference room and introduced me to the Ofsted team as the Acting Head of Science... I am not sorry to say I revealed everything to the Inspectors and was promptly sent home and threatened with dismissal from the secondment school's SLT. I never did return. The school went into special measures and I left my school after a big disagreement with the Head over acceptable and reasonable behaviour... What is the most unprofessional situation you have been placed in?

The question she raises is an interesting one. This example is also included in a TES article 'Tricks of the trade' (Stewart, 2012) which discusses the underhand tactics increasingly being used by schools to trick inspectors, according to teachers. The range of methods, as well as 'borrowing' outstanding staff from neighbouring schools to take lessons, includes telling their own weak teachers to go off sick, borrowing brilliant artwork from other schools to impress inspectors and even paying students to absent themselves. Despite having good attendance records, poorly behaved students were being paid to truant to prevent their schools getting bad ratings. This TES article (www.tes.org.com) promoted an official response from Ofsted (see Box below) and very soon after it was announced, coincidentally perhaps, that there were plans to introduce no-notice inspections for all regular school and college inspections from September 2012. Speaking about the plans, Sir Michael Wilshaw HMCI, said:

> Ofsted has been moving towards a position of unannounced school inspection over a period of years. I believe the time is now right for us to take that final step and make sure that for every school we visit inspectors are seeing schools as they really are in the corridors, classrooms and staff room. (Ofsted Press release – 10 January 2012)

Figure 7.1 Ofsted responds to reports about inspection (06 Jan 2012)

Ofsted National Director, Sue Gregory, has responded to news reports alleging that some schools are attempting to manipulate Ofsted inspections.
She said: 'In over 5,500 school inspections conducted by professional and highly trained inspectors last year there have been only a handful of issues raised with us about possible misrepresentation of the school's position, all of which were looked into. Schools have no more than two days notice of inspection and, with a thorough examination of the schools record, including attendance levels and staffing details, any sudden changes are readily observed. While we do not take suggestions of wrong doing lightly, it would be a disservice to all those schools who strive to do the best for their pupils to suggest that there is some sort of wide-scale problem based on anonymous and unsubstantiated claims'.

5.4. Explaining unintended consequences

The previous section indicated interesting differences in unintended consequences between countries and provided an example from England. These findings lead us to question the causes of unintended consequences and the specific elements of school inspection systems that motivate such strategic behaviours. A paper by Altrichter and Kemethofer (2015) and Jones et al (in prep) sheds some light on this question in explaining how "accountability pressure" on schools is associated with unintended negative effects. Altrichter and Kemethofer (2015, p.10) define such pressure as '*pressure* on individual schools and their representatives to act in conformity with the standards of an accountability system and to take action to improve school quality and effectiveness'. Pressure is expected to motivate school staff to anticipate how they are being evaluated and how targets are set, and to modify their behaviour accordingly. Unintended consequences arise when performance targets and measures are imperfect representations of the actual efforts and performance they purport to measure, which is likely to be the case when measuring something as complex as educational quality (Smith, 1995).
Altrichter and Kemethofer (2015) and Jones et al (in prep) compared the high pressure inspection systems in England and the Netherlands to the low(er) stakes systems in Austria,

Sweden, Ireland and the Czech Republic. High stakes systems are, according to these authors, systems that target failing schools for intensive monitoring, have thresholds in place to distinguish failing schools, sanction for low-performing schools, and report inspection assessments of individual schools to the general public to systems that are considered to be low stakes. Reporting school inspection reports may enhance accountability pressure as a school's stakeholders will notice and make use of inspection results, according to Maritzen (2007). Altrichter and Kemethofer (2015) found that principals in high stakes inspection systems experience more "accountability pressure" and also report of significantly more unintended consequences. School leaders in the Austrian "low-pressure accountability" system say less often that they take measures to discourage new teaching strategies and to narrow the curriculum than those in the English system, which are characterized by more accountability pressure. Schools with lower Ofsted grades, who are under pressure to improve, particularly report of narrowing the curriculum and instructional strategies in schools. These schools are under most pressure to improve quickly and apparently revert to quick fix solutions to do so. There is however limited evidence on whether these side effects disappear when school move to more positive inspection categories.

Pressure is, according to Jones et al (in prep), however not the only explanation for unintended consequences. Ehren et al (tr) for example describe how mechanisms that cause school improvement from inspections equally result in side effects. The fact that inspection systems *set expectations* for principals and teachers on good education seems to lead on the one hand to positive outcomes where schools improve the quality of teaching and the organisation of the school, but oftentimes also lead to unintended consequences when principals and teachers choose strategies that are dysfunctional in the long run, such as when they overly narrow their curriculum and instruction to meet inspection criteria.
The example presented from England allows us to understand and reflect on how this mechanism operates. The case study shows how school inspections cause schools to internalise expected behaviours, and learn these behaviours through acceptance of a discourse. Schools learn how to perform within the accepted norms of an 'effective school', a concept inspired by the discourse of school effectiveness research. Phrases such as 'experts say…', 'studies show…', and 'research has concluded…' give power to the concepts underlying inspection standards and frameworks; these concepts are incorporated in the environment in which schools and teachers operate (e.g. through curriculum frameworks, assessment, self-evaluations) and create an environment in which there is little room for schools to deviate from these concepts. As Ball (1990, p.162) explains 'teachers are trapped into taking responsibility for their own disciplining through schemes of self-appraisal, school improvement and institutional development'.

A relevant question to ask is when these processes of normalisation and setting expectations lead to genuine improvement and when they lead to unintended consequences of reinforcing strict and potentially ineffective recipes for how a school should be run. Jeffrey and Woods (1998, p.106) argue that the school effectiveness discourse as enforced by Ofsted inspectors has led to a colonisation of schools where over-mechanistic approaches of replicating effectiveness factors from research become the dominant model over time. Thus schools having learnt the accepted modes of behaviour in preparing for inspection visits, they continue to perform the good school between inspections until that becomes how the school functions all the time.

Such behaviour is reinforced through the implementation of internal quality assurance and monitoring which creates a situation of continuous surveillance where those concerned come to discipline themselves (Harland, 1996, p.101). Wilcox and Gray (1996, p.120) also link inspection with disciplinary mechanisms, 'as it requires a school to undergo an exacting discipline which extends over a period considerably longer than that of the inspection week and may also lead to a school being disciplined'. Inspection handbooks continue to influence schools in between inspections and are often used as a management development tool. As

Troman (1997, p.349) describes, 'inspectors are the absent presence in the school'. This practice is actively promoted by Ofsted as former HMCI Chris Woodhead once remarked that headteachers should be seen – and see themselves - as 'resident inspectors' (see also MacBeath, 2006).
Unintended consequences arise when accepted models of behaviour include scripted blueprints and taxonomies for what an effective school looks like, categorizing schools and teachers that have not implemented such blueprints and taxonomies as ineffective or failing, when in fact the research provides a much more fuzzy, nuanced and contextualized picture of conditions of school effectiveness as we explained in chapter 2.

This raises another explanation for the prevalence of unintended consequences, which relate to difficulties in how performance of schools is measured. As Smith (1995) explains performance measurement schemes, such as inspection frameworks, often *lack precision or fidelity in measuring complex phenomena*, such as education quality. These schemes oftentimes emphasize quantifiable aspects of performance, offering a snapshot of the school's activities and ignoring other unquantifiable, but equally important objectives and activities that may have an impact in the long term. Controllers, such as school inspectors, are often unable to process performance data correctly, such as when they have to interpret vast amounts of test results and apply complicated formulas to correct for school intake and socio-economic backgrounds of students. Such limitations in adequately measuring school performance will cause unintended strategic responses, according to Smith (1995) as schools will focus on the quantifiable aspects of their performance as measured by school inspectors (e.g. student performance on tested subjects), and ignore aspects of their service delivery that are beyond the scope of the inspection framework.

Performance measure schemes are also often *inflexible* and not equipped to respond to new circumstances or adapt to specific contexts in which schools operate. As a result, they may not capture specific circumstances in which schools function well, and not match well to school internal objectives and values, creating an overall lack of buy-in for inspections and specific inspection standards and an incentive to game and manipulate the inspection. School staff are likely to feel an incentive to manipulate inspection data when they feel inspections are 'done to' them, instead of something they actively engage in. Such inflexibility may also cause organizational paralysis when inspectors rigidly apply the inspection framework when evaluating school performance. Equally, school inspectors may neutralize some of these unintended consequences when visiting a school, such as when they contextualize their assessment, and discuss school documentation and data with school staff to understand the actual achievements of the school and establish the accuracy of the presented facts and figures.

The inspection *feedback and tone of voice* school inspectors use during their visit seems another important condition of unintended consequences. As Bates (2013) suggests, inspection feedback which focuses on negative issues and error detection may undermine confidence of school staff, decrease commitment for improvement and increase cynicism and resentment about the inspection process. The tone of voice of inspectors and whether inspectors take on an 'inspectoral role' instead of a more advisory approach during visits is an important reason for reducing trust in external inspections, and causes frustration and stress of school staff according to Berry (2012), and Brimblecombe et al (1995). De Waal (2006) describes how Ofsted inspectors intimidate schools into compliance, degrade teachers and foster fabrication of evidence to meet tick boxes, enforcing a top down dictat about how teachers should teach. As a result, teachers and head teachers in this study felt that Ofsted inspections were demoralizing and destructive and led to a climate of fear.

Credibility and training of school inspectors are important in creating more positive conditions and in ensuring a positive inspection experience for schools, particularly when school inspectors are trained to engage with schools in reflecting on improvement and

creating a safe environment in which schools are willing to share weaknesses (see Memduhoglu, 2012, Penninckx, 2015). Smith (1995) emphasizes that any type of control needs to be exercised with great care and discretion as measuring performance is a complex, dynamic and ill-defined process. Many organizational outputs, such as school quality and outcomes are the result of a joined and collaborative process which unfolds over a long period of time, highlighting the need for highly skilled inspectors to evaluate such quality and outcomes. A high trust environment, where school staff are open about their performance and reflect on improvements is an important condition for effective inspections, particularly as all performance indicators will have side effects when school staff are under pressure to perform well and eventually learn how to game and manipulate inspection indicators.

References

Altrichter, H. & Kemethofer, D. (2015) Does accountability pressure through school inspections promote school improvement? School Effectiveness and School Improvement: An International Journal of Research, Policy and Practice, 26:1, 32-56.

Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of education policy*, *18*(2), 215-228.

Bates, A. (2013). Transcending systems thinking in education reform: implications for policy-makers and school leaders. Journal of Education Policy, 28(1), 38-54.

Berry, J. (2012). Teachers' Professional Autonomy in England: Are Neo-Liberal Approaches Incontestable? *FORUM: for promoting 3-19 comprehensive education*, *54*(3), 397-409.

Booher-Jennings, J. (2005). Below the Bubble: 'Educational Triage' and the Texas Accountability System. *American Educational Research Journal*, 42(2), 231-268.

Braun, A., Maguire, M., & Ball, S. J. (2010). Policy enactments in the UK secondary school: examining policy, practice and school positioning. Journal of Education Policy, 25(4), 547-560.

Brimblecombe, N., Ormston, M., & Shaw, M. (1995). Teachers' Perceptions of School Inspection: a stressful experience. Cambridge Journal of Education, 25(1), 53-61. doi: 10.1080/0305764950250106

Brookes, C (Gen Sec of NAHT) Ch. 8 We need an inspection process. But not this one. In DE WAAL, A. (Ed) & INSTITUTE FOR THE STUDY OF CIVIL SOCIETY 2008. Inspecting the inspectorate: Ofsted under scrutiny, London, Civitas.

Campbell 1976 Campbell, D. T., "Assessing the Impact of Planned Social Change". Occasional Paper Series, Paper 8, The Public Affairs Center, Dartmouth College, Hanover New Hampshire, USA. December, 1976.

Case, P., et al. (2000). "Please Show You're Working: A Critical Assessment of the Impact of OFSTED Inspection on Primary Teachers." British Journal of Sociology of Education 21(4): 605-621.

CHAPMAN, C. 2000. Improvement, Inspection and Self-Review. *Improving Schools*, 3, 57-63.

Chapman, C. (2001). Changing Classrooms through Inspection. School Leadership & Management, 21(1), 59-73.

Chapman, C. 2002. Ofsted and School Improvement: Teachers' Perceptions of the Inspection Process in Schools Facing Challenging Circumstances. School Leadership & Management, 22, 257-272.

Courtney, S. J. (2012). Ofsted's revised school inspection framework: experiences and implications. Paper presented at BERA conference, Manchester.

Courtney, S. J. & Steven, J. C. 2013. Head teachers' experiences of school inspection under Ofsted's January 2012 framework. United Kingdom.

Cullen, J.B. and Reback, R. (2006). Tinkering toward accolades: school gaming under a performance accountability system. Cambridge: National Bureau of Economic Research. Working paper 12286. http://www.nber.org/papers/w12286.

Day, C., & Gu, Q. (2010). The new lives of teachers. London: Routledge.

Dean, J. 1995. What Teachers and Headteachers Think about Inspection. Cambridge Journal of Education, 25, 45-52.

De Waal, A. 2006 Inspection, inspection, inspection: how Ofsted crushes independent schools and independent teachers. London, The Institute for the Study of Civic Society

De Wolf, I. F., & Janssens, F. J. G. (2007). Effects and side effects of inspections and accountability in education: an overview of empirical studies. Oxford Review of Education, 33(3), 379-396.

Duffy, M. (2005, April 22). Friday forum: Inspection. Times Educational Supplement.

Duffy, B. (1999). The analysis of documentary evidence. In: J. Bell (Ed). Doing Your. Research Project: A Guide for First time Researchers in Education and Social Science (p.106-117). Buckingham: OUP.

M. Ehren., and N. Shackleton (2014) Risk-based school inspections: impact of targeted inspection approaches on Dutch secondary schools Unpublished manuscript

Ehren, M.C.M., Altrichter, H., McNamara, G. and O'Hara, J. (2013). Impact of school inspections on teaching and learning; Describing assumptions on causal mechanisms in six European countries. Educational Assessment, Evaluation and Accountability. 25(1), 3-43. http://dx.doi.org/10.1007/s11092-012-9156-4

Ehren, M.C.M. and Swanborn, M. (2012). Strategic data use of schools in accountability systems. *School Effectiveness and School Improvement 23(2),* 257-280.

Figlio, D.N. and Getzler, L.S. (2002). Accountability, ability and disability: gaming the system. *NBER working paper 9307.* http://www.nber.org/papers/w9307.

Fitz-Gibbon, C. T. (1997). The Value Added National Project: Final Report: Feasibility studies for a national system of Value Added indicators. London: School Curriculum and Assessment Authority.

Fitz-Gibbon, C. T., & Stephenson-Forster, N. J. (1999). Is OFSTED helpful. *An inspector calls: OfSTED and its effect of school standards. London: Kogan Page*.

Follows, M. 2001. After the storm the tale of an Ofsted inspection. Management in Education, 15, 24-28.

Gray, C. & Gardner, J. 1999. The Impact of School Inspections. Oxford Review of Education, 25, 455-468.

Hall, C., & Noyes, A. (2007). *The impact of school self evaluation processes on British teachers' views of their work and professionalism*. Paper presented at the BERA, London available from http://www.leeds.ac.uk/educol/documents/167834.doc

Haladyna, T.M., Nolen, S.B., Haas, N.S. (1991). Raising Standardized Achievement Test scores and the Origins of Test Score Pollution. *Educational Researcher, 20(5),* 2-7.

Hardy, I. (2012). "Managing" Managerialism: The Impact of Educational Auditing on an Academic "Specialist" School. European Educational Research Journal, 11(2), 274-289.

Harland, J. (1996). 6 Evaluation as realpolitik. *Understanding educational research*, 91.

Hopkins, D., Harris, A., Watling, R. & Beresford, J. 1999. From Inspection to School Improvement? Evaluating the Accelerated Inspection Programme in Waltham Forest. British Educational Research Journal, 25, 679-690.

Hussain, I. (2012). Subjective performance in the public sector: evidence from school inspections: London School of Economics and Political Science. Centre for Economic Performance.

Jacob, B.A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89(5-6), 761-796.

Jacob, B.A. and Levitt, S.D. (2003). Rotten Apples: an investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics (august)*, 843-877.

Jeffrey, B. & Woods, P. 1996. Feeling Deprofessionalised: the social construction of emotions during an OFSTED inspection. Cambridge Journal of Education, 26, 325-343.

Jeffrey, B., & Woods, P. (1998). *Testing teachers: The effect of school inspections on primary teachers*. Psychology Press.

Jones et al (submitted). The unintended consequences of school inspection: The prevalence of inspection side-effects in Austria, the Czech Republic, England, Ireland, the Netherlands, Sweden and Switzerland.

Keeble-Allen, D. 2004. Inspection at Summerhill: did Ofsted inspection result in improvement? U210580 Ed.D., University of Leicester (United Kingdom).

Kogan, M. & Brunel University, Centre For The Evaluation Of Public Policy And Practice 1999. The Ofsted system of school inspection: an independent evaluation, Uxbridge, Brunel University Centre for the Evaluation of Public Policy and Practice.

Koretz, D.M., McCaffrey, D.F. and Hamilton, L.S. (2001). Towards a Framework for Validating Gains under High-Stakes Conditions. CRESST/Harvard Graduate School of Education: CSE Technical Report 551

Lee, J. & Davies, J. D. 2000. Bereavement and Loss: the Emotional Consequences of Special Measures for a Sample of Special Schools. Improving Schools, 3, 44-50. (not in spreadsheet as based on same data set as reference below)

Lupton, R., & Hempel-Jorgensen, A. (2012). The importance of teaching: pedagogical constraints and possibilities in working-class schools. Journal of Education Policy, 27(5), 601-620.

MacBeath, J. (2008) Leading learning in the self-evaluating school. *School leadership and management: formerly school organisation*, 28: 4, 385-399

McNamara, G., O'Hara, J., Lisi, P. L., & Davidsdottir, S. (2011). Operationalising self-evaluation in schools: experiences from Ireland and Iceland. *Irish Educational Studies, 30*(1), 63-82.

Memduhoglu, H. B. (2012). The Issue of Education Supervision in Turkey in the Views of Teachers, Administrators, Supervisors and Lecturers. Educational Sciences: Theory and Practice, 12(1), 149-156.

Merton, R. K. 1936. "The unanticipated consequences of purposive social action."*American Sociological Review* 1: 894-904.

Nees, P. (2006). SCHOOLS AND THEIR E.R.O. RECOMMENDATIONS: a study of six Wellington area schools – May / June 2006 Sabbatical Report.

Ouston, J., et al. (1997). "What Do Schools Do after OFSTED School Inspections--Or Before?" School Leadership & Management 17(1): 95-104.

Penninckx, M. (2015) Inspecting School Inspections. Doctoral dissertation. University of Antwerp

Perryman, J. 2002. Surviving Special Measures: A Case Study of a 'Fresh Start' School. Improving Schools, 5, 46-59.

Perryman, J. (2006). Panoptic Performativity and School Inspection Regimes: Disciplinary Mechanisms and Life under Special Measures. Journal of Education Policy, 21(2), 147-161.

Perryman, J. (2007). Inspection and Emotion. Cambridge Journal of Education, 37(2), 173-190.

Perryman, J. (2009). Inspection and the Fabrication of Professional and Performative Processes. Journal of Education Policy, 24(5), 611-631.

Plowright, D. (2007). "Self-evaluation and Ofsted Inspection Developing an Integrative Model of School Improvement." *Educational Management Administration & Leadership 3*5(3): 373-393.

ROBERTS-HOLMES, G. 2014. The 'datafication' of early years pedagogy: 'if the teaching is good, the data should be good and if there's bad teaching, there is bad data'. *Journal of Education Policy*, 1-13.

Scanlon, M. (1999). The impact of OFSTED inspections, London, National Foundation for Educational Research for the National Union of Teachers.

Sebba, J., Clarke, J. & Emery, B. 1996. How can the inspection process enhance improvement in special schools? European Journal of Special Needs Education, 11, 82-94.

Smith, P. 1995. "On the unintended consequences of publishing performance data in the public sector." *International Journal of Public Administratio*n 18(2 and 3): 277-310.

Stecher, B.M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practices). Tests and their use in test-based accountability systems. In Hamilton, L.S., Stecher, B.M., Klein, S.P. (Eds.). *Making sense of Test-based Accountability in Education*. Santa Monica: Rand cooperation. http://www.rand.org/pubs/monograph_reports/MR1554/

Sturman, L. (2011). Test Preparation: Valid and Valuable, or Wasteful? *The Journal of the Imagination in Language Learning, 9*, 31-31

Troman, G. (1997). Self-management and school inspection: complementary forms of surveillance and control in the primary school. *Oxford review of education, 23*(3), 345-364.

Tymms, P. 2004. "Are standards rising in English primary schools?" *British Educational Research Journal* 30 (4) : 477-494.

Webb, R., Vulliamy, G., Häkkinen, K. & Hämäläinen, S. (1998). External Inspection or School Self-evaluation? A Comparative Analysis of Policy and Practice in Primary Schools in England and Finland. *British Educational Research Journal*, 24, 539-556.

West, A., Mattei, P., & Roberts, J. (2011). Accountability and Sanctions in English Schools. *British Journal of Educational Studies, 59*(1), 41-62.

Wilcox, B. & Gray, J. 1994. Reactions to Inspection: a study of three variants. Cambridge Journal of Education, 24, 245-269.

Wilcox, B., & Gray, J. (1996). *Inspecting Schools: Holding Schools to Account and Helping Schools To Improve*. Buckingham: OUP

Wiggins, A. and Tymms, P. 2002. "Dysfunctional Effects of League Tables: A Comparison between English and Scottish Primary Schools." *Public Money and Management* 22(1), 43-48.

Wong, M. N., & Li, H. (2010). From External Inspection to Self-Evaluation: A Study of Quality Assurance in Hong Kong Kindergartens. Early Education and Development, 21(2), 205-233.